

# Exploring German Secondary School Students' Conceptual Knowledge of Density

Tim Zenger\*, Philipp Bitzenbauer

Physics Education, Department of Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

\*Corresponding Author: [tim.zenger@fau.de](mailto:tim.zenger@fau.de)

## ABSTRACT

This article reports on the development and piloting of a German version of a concept test to assess students' conceptual knowledge of density. The concept test was administered in paper-pencil format to 222 German secondary school students as a post-test after instruction in all relevant concepts of density. We provide a psychometric characterization of the instrument based on the student responses. Therefore, we use methods from classical test theory and item response theory. Our analysis shows that our concept test has good psychometric properties. Furthermore, it is shown to enable a reliable ( $\alpha=0.77$ ) survey of students' conceptual knowledge of density. Finally, we use the student responses to identify secondary school students' difficulties with the density concept. We replicate student difficulties on density from the literature, but our study also reveals some new difficulties.

**KEY WORDS:** Concept inventory; student difficulties; density

## INTRODUCTION

Density is an integral part of introductory physics courses, not only in German secondary schools. In lessons on density, learners make their first experiences with diverse aspects related to the discipline of physics: "First introduced in elementary grades with the ideas of floating and sinking, density taught in middle school is geared toward understanding through the use of mathematical formulas" (Dawkins et al., 2008, p. 21). For example, the new nine-year Gymnasium (secondary school) curriculum, which has only recently entered into force in Bavaria, a state in the south of Germany, includes density as an introductory topic (cf. ISB, 2021). Specifically, in their first physics lessons, the learners in the seventh grade (12–13 years) are supposed to learn about numerous physics methods and experience Nature of Science aspects in addition to the subject content. For instance, it is envisaged that the students (ISB, 2021).

- ...plan an experiment to determine the density of an irregularly shaped body, which they carry out and protocol under guidance.
- ...identify the material of the body using tables from the internet, taking into account the limited accuracy of measured values.
- ...perform simple calculations on density and mass, taking into account units and the tolerance of measurements.
- ...identify the density of materials as a physical criterion of judgment in texts on modern technology and distinguish it from non-specialist criteria.

Given the importance of density in early physics education in German secondary schools (cf. Lichtfeldt and Peuckert, 1997;

Uhden, 2016), developing an instrument that allows teachers to survey their students' conceptual knowledge of density economically seems necessary. This article reports on the development and piloting of a German version of the *Density Survey*, a concept test to assess students' conceptual knowledge of density originally developed by Yeend et al. (2001) in the English language. We provide a psychometric characterization, focusing on item statistics, test reliability, and Rasch scaling: Consequently, we use methods from classical test theory and item response theory in this research. Finally, we use the results from 222 students to identify German secondary school students' difficulties with the density concept.

## LITERATURE REVIEW

### Conceptual Knowledge and Concept Inventories

Conceptual knowledge "consists of the core concepts for a domain" and "has been characterized using several different constructs including semantic nets, hierarchies, and mental models" (Byrnes and Wasik, 1991, p. 777). Assessing learners' conceptual knowledge on a given topic can help teachers understand how learners think so that they can adapt and adjust their teaching accordingly (cf. Furrow and Hsu, 2019). Concept inventories provide a way of surveying students' conceptual knowledge on a given topic. A concept inventory is an "instrument designed to evaluate whether a person has an accurate and working knowledge of a concept or concepts" (Lindell et al., 2007, p. 1) – mainly using single or multiple-choice items, respectively. Hence, they offer a particular advantage regarding their usage in schools, for example, compared to interviews: Their implementation and evaluation

are significantly less time-consuming.

Research-validated concept inventories have the potential to be of value for evaluating the effectiveness of a particular curriculum or pedagogy in order to ensure that students grasp the fundamental concepts of a specific topic. Furthermore, concept inventories have been used to identify students' conceptions of different areas in science education research (cf. Treagust, 1988), such as mechanics (Hestenes et al., 1992; Hestenes and Halloun, 1995), electricity (Engelhardt and Beichner, 2004), magnetism (Li and Singh, 2016), optics (Kaltakci-Gurel et al., 2017), thermodynamics (Chu et al., 2012), or quantum physics (McKagan et al., 2010).

The quality of concept inventories' single or multiple-choice items largely depends on the distractors' quality. Distractors are those answer options of items that are false (D'Avanzo, 2008), and Moosbrugger and Kelava (2012) define distractors as "answer alternatives that seem plausible but do not apply" (p. 418). Hence, distractors are usually based on widespread student conceptions (Bitzenbauer, 2021). That is why we provide a brief overview of students' conceptions on density in the next section.

### Students' Conceptions of Density

Conceptual knowledge of density requires students' understanding of ratios (Rowell and Dawson, 1977) and proportional relationships (Yeend et al., 2001), which is why Inhelder and Piaget (1958) used density tasks in their well-known work to assess reasoning ability. Numerous empirical studies have explored student conceptions of density (Dawkins et al., 2008; Smith et al., 1997). It was shown that density as a material-specific quantity is abstract for learners (Kang et al., 2004), so an improper equation of density with volume, mass, or size is often observed (Smith et al., 1986; Yeend et al., 2001). For example, Stepan et al. (1986) found that "a large number of college students said that when you crumple an aluminum sheet, you have made it easier" (p. 66). A mix-up of density with the term weight used in everyday language has also been observed among learners (Driver et al., 2014). In their article, Xu and Clarke (2012) observe:

- Microscopic conceptions of density, for example, density "measures how close particles are to each other (p. 786) which are sometimes implicitly juxtaposed with
- Macroscopic conceptions in students' thinking, for example, density of an object explains why it will float or sink in water" (p. 786).

Such conceptions can, of course, be the starting point for many further learning difficulties. For instance, many students believe that heavier objects sink more. Researchers argue that an improper equation of density with volume, mass, or size may be the origin of such conceptions (cf. Potvin et al., 2015). Similar difficulties regarding the density concept have recently been found among science teacher candidates using a four-tier diagnostic instrument (Kiray and Simsek, 2021).

## OBJECTIVES OF THIS STUDY

The research objectives of this paper are threefold:

1. We aim at providing a concept test to assess German secondary school students' conceptual knowledge of density economically in classroom scenarios.
2. We target a psychometric characterization of the instrument, especially concerning discriminatory power and reliability.
3. We explore difficulties German secondary school students have with concepts of density that we found from student responses.

## DESIGN OF A GERMAN VERSION OF THE DENSITY SURVEY

Yeend et al. (2001) developed the *Density Survey*, a concept inventory to examine student understanding of density. The development was based on results from a literature review as well as interviews with high school students, precollege science teachers, and university lecturers (Yeend et al., 2001). The *Density Survey* comprises 11 single-choice items for evaluating students' conceptions of six key aspects of density (one or two items per key aspect), identified through expert surveys (cf. Table 1). Among the 11 items, one on proportional reasoning was taken from the TIMSS study (Trends in International Mathematics and Science Study). In addition, three items were adapted from a concept inventory on electric charge density (Yeend et al., 2001). The instrument was piloted with 787 students of different ages and school types in the USA (cf. Yeend et al., 2001).

To develop a concept test on density for German-speaking countries, Yeend et al.'s (2001) *Density Survey* was first translated into the German language. The items were then reviewed and revised in German according to the guidelines for creating multiple-choice tasks according to the Haladyna et al. (2002). Our German adaption can be accessed through PhysPort<sup>1</sup>. For sample items, we refer to the PhysPort link. However, an overview of the test instrument items' contents is given in Table 1. The availability of our German version of the *Density Survey* on the PhysPort platform ensures free access for interested teachers who want to use the instrument in their courses.

## RESEARCH DESIGN

### Administration of the Survey

The final version of our German version of the *Density Survey* was administered in paper-pencil format to 222 German secondary school students (Grade 7, students aged 12–13) from three public secondary schools as a post-test after instruction in all relevant concepts of density (test duration: 10–15 minutes). The students' teachers administered the test. All students participated in our study voluntarily, according to the ethical

1 <https://www.physport.org/assessments/assessment.cfm?I=54&A=DS>

standards (cf. *ethical statement*). After test administration, we used the student responses for statistical analysis to obtain information about the psychometric quality of the instrument. Data analysis methods are described in the following.

### Psychometric Characterization of the Instrument

For the psychometric characterization of the instrument in the sense of classical test theory, we used the student responses to calculate item difficulty indices, discrimination indices, and Cronbach's alpha as a measure of internal consistency (Taber, 2018). For the item difficulty indices, we refer to the accepted range of 0.2–0.8 according to Kline (2015), and for the discriminatory indices, values above 0.2 are considered acceptable (Jorion et al., 2015). For the Cronbach alpha coefficient, we refer to the lower threshold value of 0.7, according to the Streiner (2003). Table 2 provides an overview of these indices' calculations.

### Rasch Scaling

In addition to classical test theory, we also used item response theory (IRT) to evaluate our German version of the *Density Survey*. Rasch scaling of the data was carried out using a dichotomous Rasch model (Rasch, 1960) because neither a logical dependency of the items nor a uniform item difficulty was assumed (Rost, 2004). Furthermore, we used the R-package *sirt*<sup>2</sup>, to confirm essential unidimensionality of the

2 <https://cran.r-project.org/web/packages/sirt/sirt.pdf>

Item No.	Content aspect
1	Relationship volume – size
2	Density as a material property
3	Mass
4	Density-volume dependence, $m=\text{const.}$
5	Density as a material property
6	Proportional reasoning
7	Density-mass dependence, $V=\text{const.}$
8	Conservation of mass
9	Relationship volume – size
10	Temperature dependence of the density
11	Calculation of density of a homogenous body

test instrument finding indices  $\text{DETECT} < 0.20$ ,  $\text{ASSI} < 0.25$ , and  $\text{RATIO} < 0.36$  (Jang and Roussos, 2007).

Rasch scaling enables the person's abilities and item difficulties to be represented on a common scale. In this context, item difficulty means the level on the ability scale required to solve the item with a probability of 50%. The likelihood of solving an item as a function of the person's ability is graphically displayed with the help of item characteristic curves (Wu and Adams, 2007). In our study, marginal maximum likelihood (MML) estimation was performed to obtain the item parameters independently of person parameters.

The instrument's reliability is estimated through EAP/PV reliability and WLE reliability. Their values lay between 0 and 1 and can be compared with Cronbach's alpha from classical test theory (Rost, 2004). The person's abilities and the test's item difficulties are represented with the help of a Wright Map. Here, the person abilities are described via a histogram, with the mean person ability representing the zero point on the logit scale. The item parameters, also given in logits, are arranged according to their difficulty. Finally, Rasch homogeneity of the items was checked using the criterion  $0.5 < \text{Infit/Outfit-MNSQ} < 1.5$  according to Linacre (2002). We used the R-package TAM for Rasch scaling in version 3.6-45.

Validity issues were not further investigated in the context of our study because the development of the original version of the *Density Survey* (Yeend et al., 2001) has already gone through various validation steps.

### Investigation of Student Difficulties with Density Concepts

We added a second response tier to each item to elicit student difficulties. In this second tier, we asked respondents to indicate their certainty of response in tier one on a 5-point Likert scale (1 = I guessed, ..., 4 = sure, and 5 = very sure). Incorrect answers where learners state that they were *sure* or *very sure* indicate the presence of student difficulties (cf. Hasan et al., 1999).

## RESULTS

### Results of the Psychometric Characterization of the German Version of the Density Survey

The results of an item analysis (Table 3) show that the item difficulty indices, except for two outliers (items 1 and 3), are

Index	Description	Calculation
Item difficulty index $P$	The ratio of the number $N_c$ of correct responses on one item to the total number $N$ of students who worked on the item.	$P = \frac{N_c}{N}$
Item discrimination index $D$	For this calculation, the whole sample is split into two groups, one of which comprises students with test scores higher than the median total score of the sample (H group) and one including students with lower scores (L group). The number of correct responses on a given item in both groups is referred to as $N_H$ and $N_L$ , respectively. $N$ is the total number of students who worked on the item.	$D = \frac{N_H - N_L}{N / 2}$
Cronbach's $\alpha$ (standardized)	The number of items is referred to as $k$ , and the average intercorrelation between these items is $r$ .	$\alpha = \frac{k \cdot r}{1 + (k - 1)r}$

within the accepted range between 0.2 and 0.8. Values above the acceptable range of 0.2 are observed for the discrimination indices. Only the discrimination index of item 11 lies well below this threshold (cf. Table 3). Consequently, item 11 has been deleted from the item set for further study. The need for this is also supported by the results of the Rasch scaling (cf. next section).

The value of Cronbach's alpha as an estimator for the internal consistency of the test instrument is found to be 0.77. This value can be raised by excluding item 11 mentioned above. For content reasons, the unique role of item 11 seems plausible since it is the only item in which the calculation of the density of a homogeneous body was asked. In contrast, the other items only addressed qualitative understanding of density.

Each of the 11 test items was coded dichotomously so that a point was awarded precisely when the correct answer option was selected. As a result, the students reached a mean score of  $m = 6.23$  ( $SD = 2.68$ ), ranging from 1 point to 11 points (cf. Figure 1). The median score was 6 points.

### Results of Rasch Scaling

The EAP/PV reliability of 0.79 and the WLE reliability of 0.73 of our German version of the *Density Survey* calculated from the responses of 222 German secondary school students are above the threshold accepted in the literature (Field, 2009).

The analysis of the fit parameters of all items shows that one item lacks sufficient Rasch conformity (item 11: Infit-MNSQ 1.54, Outfit-MNSQ 2.71). This item asked about calculating the density of a homogeneous body (cf. Table 1). Item 11 was therefore excluded from the item set after our analysis. Furthermore, for two items, we found Infit-MNSQ 0.72 and Outfit-MNSQ 0.23. However, we assume acceptable conformity to the Rasch model for both items: On the one hand, the T-value for Infit-MNSQ is  $-0.64$  for both items, thus, within the acceptable range of  $|T| < 2$  (Wu et al., 2007). On the other hand, the T-value for Outfit-MNSQ is  $-2.66$  for both items indicating overfit, which is regarded as unproblematic (Jude, 2006). For the remaining items, outfit statistics ranging from 0.59 (Item 8) to 1.58 (Item 7) and infit statistics ranging from 0.74 (Item 10) to 1.27 (Item 7) were found.

Table 3: Item analysis results in the overview			
Item	Difficulty index	Discriminatory index	$\alpha$ if item deleted
1	0.94	0.42	0.76
2	0.32	0.49	0.74
3	0.94	0.42	0.76
4	0.44	0.53	0.73
5	0.56	0.68	0.71
6	0.57	0.35	0.76
7	0.67	0.21	0.77
8	0.21	0.56	0.73
9	0.51	0.40	0.75
10	0.40	0.68	0.71
11	0.67	0.03	0.80

The Wright-Map analysis confirms the assumption that different items have different difficulty levels (cf. Figure 2). The person abilities range from around  $-4$  to around  $3$  logits, while the item difficulties do not exceed  $1.77$  logits. Hence, at the upper end of the scale (person ability  $> 2.0$  logits), there is a lack of suitable items to optimally record and differentiate between persons with different levels of competence. In addition, some tasks seem to capture the same area of person ability concerning the latent variable (e.g., items 5, 6, and 9). The mean item difficulty is  $-0.61$  logits reflecting that (i) higher performance ranges are slightly dominant and that (ii) individual items tend to be too easy. However, a substantial match between item difficulties and person abilities may be observed.

In Figure 3, the item-characteristic curves for all eleven items of the German version of the *Density Survey* are shown.

### German Secondary School Students' Difficulties with Density Concepts

Using incorrect answers where the respondents stated they were *sure* or *very sure*, we explored student difficulties that German secondary school students have with aspects of

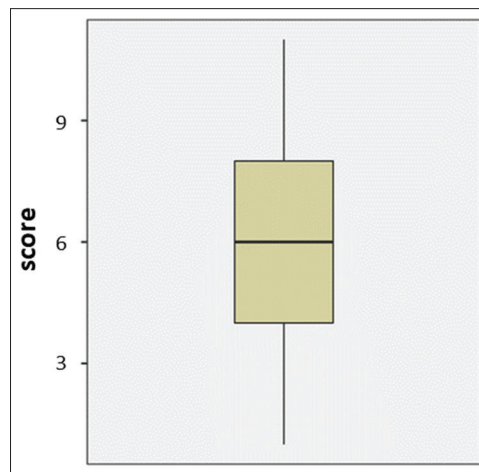


Figure 1: Box plot showing the distribution of students' test scores

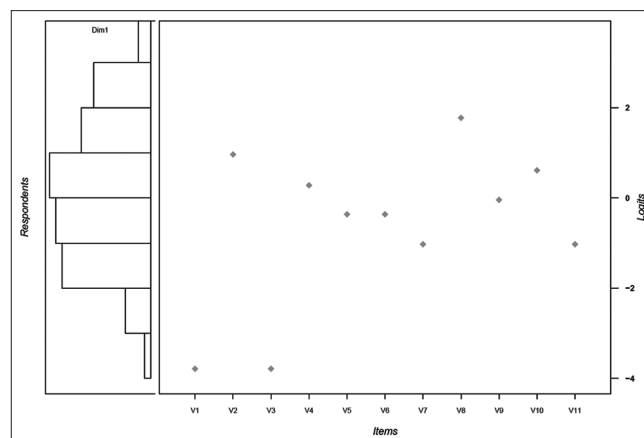
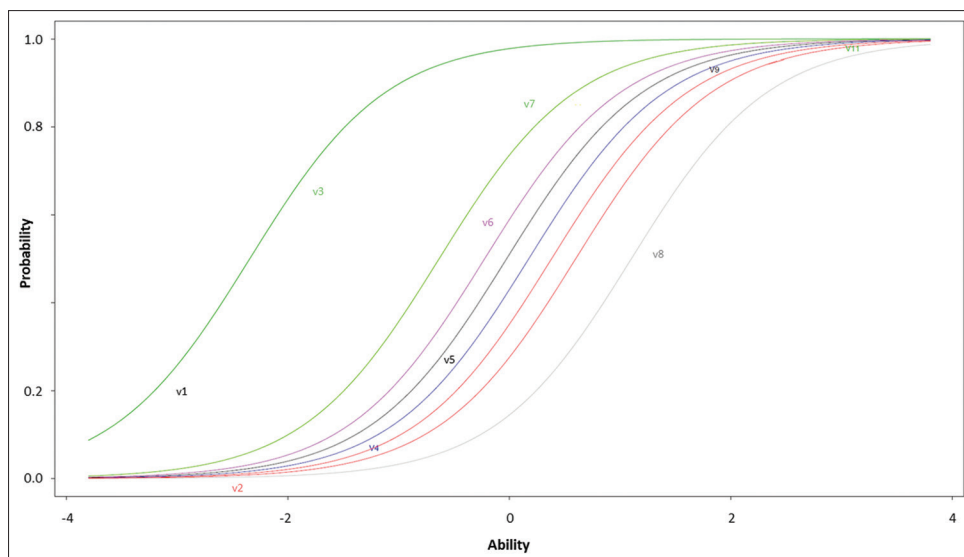


Figure 2: Wright-Map



**Figure 3:** Item-characteristic curves for all items

density. This process made it possible to replicate student difficulties on density from the literature (cf. Kang et al., 2004; Smith et al., 1986; Stepan, 1986).

A general difficulty that we observed is that students mix up quantities such as mass, volume, or density with the size of a given object. This student difficulty has already been reported by Yeend et al. (2001). For example, Item 2 refers to three identical width and height pieces cut from a straight, uniform plank piece (homogeneous). Here, 39% of the respondents were sure that the density of the largest piece (in length) would be greatest, while 17% were sure that the smallest piece would have the greatest density. A similar observation can be drawn from the responses to Item 5: In this item, the students are confronted with a jeweler cutting a small piece from a large uncut diamond. Students were then asked to compare the density of the original diamond to that of the small cut piece. Again 22% of the students answered that the density of the small piece would be smaller than the density of the original diamond and indicated that they were very sure that their answer was correct.

In Item 8, students were first told that a balloon taken outside on a freezing day would shrink even though no air escapes from the balloon. They were then asked to compare the air mass in the balloon before and after shrinkage. In our study, 61% of the students stated that the mass of the air would decrease and reported that they were (very) sure that their answer was correct. Hence, nearly two-thirds did not seem to be aware of the conservation of mass. In Item 9, even 22% answered that the air volume inside the balloon would be the same after shrinking compared to before.

Consequently, 33% of the students also thought that the air density in the balloon after shrinking was smaller than before (Item 10), while 5% answered that the density would be constant. This supports that several student difficulties

with density concepts originate from a mix of quantities such as mass, volume, or density with the size of a given object. From the results for the last-mentioned Item 10, it can also be deduced that a proportion of the students do not seem to be aware of the temperature dependence of air density. Stavy et al. (2002) argue that many learners' conceptions are based on the *More A-More B principle*, which is also reflected by the results of our study with German secondary school students.

## DISCUSSION AND CONCLUSION

This article reports on a translation and adaptation of the *Density Survey*, a concept inventory on density, into the German language. The aim was to enable teachers to survey their students' conceptual knowledge of density in a time-efficient way because density plays an essential role in introductory physics education in Germany, as shown in this paper's introduction.

A pilot study with 222 German secondary school students revealed that the German version of the instrument has acceptable to good psychometric properties and allows for a reliable assessment of learners' conceptual knowledge about density. A Rasch analysis confirmed these findings, and the Wright-Map shows that substantial correspondences between item difficulties and person abilities can be observed. Nevertheless, the future research needs to address a revision of some items (e.g., Items 1 and 3). Furthermore, the inclusion of additional items at the upper end of the scale should be considered to be able to distinguish between persons with different levels of competence, even in the area of high person abilities. However, the creation of additional items must always be considered carefully because the time-efficient applicability of the assessment in classroom scenarios must be maintained.

The results of our pilot study presented in this article can serve as a baseline for traditional introductory physics courses in

high schools in German-speaking countries. Teachers who cover the same concepts of density in their courses with an innovative teaching sequence or pedagogy can compare their own students' performance with those presented in this article to assess the effectiveness of their teaching approach. In addition, we have shown how our German version of the *Density Survey* can be used to collect student difficulties. The student difficulties that we found may inform physics education in Germany and beyond, and we provide a recommendation for classroom practice below.

We analyzed how confident the students were in their answer selections on the different test items to collect student difficulties. Incorrect answers where learners state that they were *sure* or *very sure* indicate the presence of student difficulties. We were able to replicate numerous difficulties found in the previous studies. However, we conclude by focusing on a particular problem that we uncovered: We found that a subset of learners perceives density as a constant quantity that is not temperature-dependent. Based on this result, we recommend that temperature dependence be addressed explicitly from the very beginning when it comes to introducing density in the classroom: In the case of gases, the temperature dependence of density is immediately obtained from Gay-Lussac's law (i.e., for constant pressure), and from our point of view introducing the density anomaly of water represents a further possibility to circumvent the preconception that density would not depend on temperature.

## ETHICAL STATEMENT

The study was conducted in accordance with ethical policy: Authors acknowledge that the research was conducted anonymously, that consent was obtained from all participants, and that all participants are informed about the publication of the results of this study.

## REFERENCES

- Bitzenbauer, P. (2021). Development of a test instrument to investigate secondary school students' declarative knowledge of quantum optics. *European Journal of Science and Mathematics Education*, 9(3), 57-79.
- Byrnes, J.P., & Wasik, B.A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, 27(5), 777-786.
- Chu, H.E., Treagust, D., Yeo, S., & Zadnik, M.G. (2012). Evaluation of students' understanding of thermal concepts in everyday contexts. *International Journal of Science Education*, 34(10), 1-26.
- D'Avanzo, C. (2008). Biology concept inventories: Overview, status, and next steps. *BioScience*, 58(11), 1079-1085.
- Dawkins, K.R., Dickerson, D.L., McKinney, S.E., & Butler, S. (2008). Teaching density to middle school students: Preservice science teachers' content knowledge and pedagogical practices. *The Clearing House*, 82(1), 21-26.
- Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics-Physics Education Research*, 2, 010105.
- Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (2014). *Making Sense of Secondary Science: Research into Children's Ideas*. 2<sup>nd</sup> ed. London, United Kingdom: Routledge.
- Engelhardt, P.V., & Beichner, R.J. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics*, 72(1), 98-115.
- Field, A. (2009). *Discovering Statistics Using SPSS*. 3<sup>rd</sup> ed. United States: Sage.
- Furrow, R.E., & Hsu, J. L. (2019). Concept inventories as a resource for teaching evaluation. *Evolution: Education and Outreach*, 12(2), 2.
- Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Hasan, S., Bagayoko, D., & Kelley, E.L. (1999). Misconceptions and the certainty of response index (CRI). *Physics Education*, 34(5), 294-299.
- Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory. A response to huffman and heller. *The Physics Teacher*, 33, 502-506.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher*, 30, 141-158.
- Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York, United States: Basic Books.
- ISB. (2021). *Natur und Technik 7*. [Nature and Technology 7]. Available from: [https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/7/nt\\_gym](https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/7/nt_gym) [Last accessed on 2022 Feb 01].
- Jang, E.E., & Roussos, L. (2007). An Investigation into the Dimensionality of TOEFL Using Conditional Covariance-Based Nonparametric Approach. *Journal of Educational Measurement*, 44(1), 1-21.
- Jorion, N., Gane, B.D., James, K., Schroeder, L., DiBello, L. V., & Pellegrino, J. W. (2015). An analytic framework for evaluating the validity of concept inventory claims. *Journal of Engineering Education*, 104(4), 454-496.
- Jude, N. (2006). IRT-Skalierung Skalierung mit ConQuest. [IRT Scaling Scaling with ConQuest.] Available from: [https://www.media.metrik.de/uploads/incoming/pub/Literatur/Folien\\_Jude+anleitung%20zu%20ConQuest.pdf](https://www.media.metrik.de/uploads/incoming/pub/Literatur/Folien_Jude+anleitung%20zu%20ConQuest.pdf) [Last accessed on 2022 Feb 01].
- Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Research in Science and Technological Education*, 35(2), 238.
- Kang, S., Scharmann, L.C., & Noh, T. (2004). Reexamining the role of cognitive conflict in science concept learning. *Research in Science Education*, 34, 71-96.
- Kiray, S.A., & Simsek, S. (2021). Determination and evaluation of the science teacher candidates' misconceptions about density by using four-tier diagnostic test. *International Journal of Science and Mathematics Education*, 19, 935-955.
- Kline, J.B. (2005). *Psychological Testing. A Practical Approach to Design and Evaluation*. United States: Sage.
- Li, J., & Singh, C. (2016). Developing and validating a conceptual survey to assess introductory physics students' understanding of magnetism. *European Journal of Physics*, 38(2), 25.
- Lichtfeldt, M., & Peuckert, J. (1997). Die behandlung der dichte im unterricht. *Naturwissenschaften im unterricht*. [The treatment of density in the classroom. Science in the classroom]. *Naturwissenschaften im Unterricht*, 8(45), 22-26.
- Linacre, J.M. (2002). What do Infit and Outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16, 878.
- Lindell, R.S., Peak, E., & Foster, T.M. (2007). Are they all created equal? A comparison of different concept inventory development methodologies. *AIP Conference Proceedings*, 883, 14.
- McKagan, S.B., Perkins, K.K., & Wieman, C.E. (2010). Design and validation of the quantum mechanics conceptual survey. *physical Review Physics Education Research*, 6(2), 020121.
- Moosbrugger, H., & Kelava, A. (2012). Testtheorie und Fragebogenkonstruktion. [Test Theory and Questionnaire Construction]. Berlin, Germany: Springer.
- Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education*, 13, 21-43.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Illinois, United States: The University of Chicago Press.

- Rost, J. (2004). *Lehrbuch Testtheorie-Testkonstruktion*. [Textbook Test Theory-Test Construction]. Germany: Hans Huber.
- Rowell, J.A., & Dawson, C.J. (1977). Teaching about floating and sinking: An attempt to link cognitive psychology with classroom practice. *Science Education*, 61(2), 245-253.
- Smith, C., Carey, S., & Wisner, M. (1986). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21, 177-237.
- Smith, C., Maclin, D., Grosslight, L., & Davis, H. (1997). Teaching for understanding: A study of students' pre-instruction theories of matter and a comparison of the effectiveness of two approaches to teaching about matter and density. *Cognition and Instruction*, 15(3), 317-393.
- Stavy, R., Tsamir, P., & Tirosh, D. (2002). Intuitive Rules: The Case of "More A-More B". In: Limón, M., & Mason, L., (Eds.), *Reconsidering Conceptual Change: Issues in Theory and Practice*. Berlin, Germany: Springer. pp. 217-231.
- Stepans, J.L., Beiswenger, R.E., & Dyché, S. (1986). Misconceptions die hard. *The Science Teacher*, 53(6), 65-69. Available from: <https://www.jstor.org/stable/24140038> [Last accessed on 2022 Feb 01].
- Streiner, D.L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99-103.
- Taber, K.S. (2018). The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273-1296.
- Treagust, D. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10, 159-169.
- Uhden, O. (2016). Warum bildet man Verhältnisse? Eine Erarbeitung von Verhältnisgrößen am Beispiel der Dichte. [Why do you form proportions? An elaboration of ratio variables using the example of density. Science in the classroom]. *Naturwissenschaften im Unterricht [Science in Class]*, 27(153/154), 28-31.
- Wasserman, J.D., & Bracken, B.A. (2003). Psychometric characteristics of assessment procedures. In: Weiner, I.B., (Ed.), *Handbook of Psychology*. New Jersey, United States: John Wiley & Sons, Inc.
- Wu, M., & Adams, R. (2007). *Applying the Rasch Model to Psycho-Social Measurement: A Practical Approach*. Tamil Nadu: Educational Measurement Solutions.
- Wu, M.L., Adams, R.J., Wilson, M.R., & Haldane, S.A. (2007). *ACER ConQuest: Version 2.0. Generalised Item Response Modelling Software*. India: ACER Press.
- Xu, L., & Clarke, D. (2012). Student difficulties in learning density: A distributed cognition perspective. *Research in Science Education*, 42, 769-789.
- Yeend, R., Loverude, M., & Gonzalez, B. (2001). Student understanding of density: A cross-age investigation. In: *Paper presented at Physics Education Research Conference 2001, Rochester, New York*.